ELSEVIER

Comparative Biochemistry and Physiology Part B 131 (2002) 453–463

CBP

www.elsevier.com/locate/cbpc

Partial characterization of trypsin-like protease and molecular cloning of a trypsin-like precursor cDNA in salivary glands of *Lygus lineolaris*

F. Zeng[a,*], Y. Zhu[b], A. Cohen[a]

[a]*Biological Control and Mass Rearing Research Unit, USDA/ARS, P.O. Box 5367, Mississippi State, MS 39762, USA*
[b]*Southern Insect Management Research Unit, USDA-ARS-JWDSRC, P.O. Box 346, Stoneville, MS 38776, USA*

Received 6 September 2001; received in revised form 16 November 2001; accepted 22 November 2001

Abstract

Based on substrate specificity, an alkaline pH optimum, sensitivity to selected proteinase inhibitors, and molecular analysis, we provide evidence for the presence of a trypsin-like serine proteinase in the salivary gland complex (SGC) of the tarnished plant bug, *Lygus lineolaris* (Palisot de Beauvois) (Heteroptera: Miridae). The predominant activity in extracts of the SGC against N_2-benzoyl-L-arginine-*p*-nitroanilide (L-BA*p*NA) was at pH 10, but a minor peak of activity also occurred at pH 5. The major BA*p*NAase activity focused at 10.4 during preparative isoelectric focusing and was eluted with an apparent molecular weight of 23,000 from a calibrated gel filtration column. The BA*p*NAase fraction gave a single major band when analyzed on a casein zymogram. The activity was completely suppressed by the serine protease inhibitors, phenylmethylsulfonyl fluoride (PMSF) and lima bean trypsin inhibitor. A cDNA coding for a trypsin-like protein in the salivary glands of *L. lineolaris* was cloned and sequenced. The 971bp cDNA contained an 873-nucleotide open reading frame encoding a 291-amino acid trypsin precursor. The encoded protein included amino acid sequence motifs that are conserved with four homologous serine proteases from other insects. Typical features of the putative trypsin-like protein from *L. lineolaris* included the serine protease active site (His^{89}, Asp^{139}, Ser^{229}), conserved cysteine residues for disulfide bridges, the residues (Asp^{223}, Gly^{252}, Gly^{262}) that determine trypsin specificity, and both zymogen signal and activation peptides. Cloning and sequencing of a trypsin-like precursor cDNA provided additional direct evidence for trypsin like enzymes in the salivary glands of *L. lineolaris.*

Keywords: Hemiptera; pH optimum; Zymogram gel; Serine protease inhibitor; Trypsin; cDNA; Serine proteinases

1. Introduction

The salivary gland complex is an important part of the feeding apparatus in extra-oral digestion (Cohen, 1998), and in hemipteran insects is reported to express proteolytic activity (Goodchild, 1952; Hori, 1970; Laurema et al., 1985; Cohen, 1993 Agusti and Cohen, 2000; Zeng and Cohen, 2000; Colebatch et al., 2001). Trypsins and/or chymotrypsins are major endopeptidases in most insects (Chapman, 1998). Trypsin-like proteases have been detected in several insect orders including Coleoptera (e.g. Gooding and Huang, 1969; Baker 1981; Zhu and Baker, 1999), Diptera (Gooding and Rolseth, 1976; Terra and Ferreira, 1983; Lemos and Terra, 1992), Hymenoptera (Giebel et al., 1971; Schumaker et al., 1993), Lepidoptera (Miller et al., 1974; Milne and Kaplan, 1993;

*Corresponding author. Tel.: +1-662-320-7530; fax: +1-662-320-7571.
E-mail address: fzeng@bcmrru.msstate.edu (F. Zeng).

1096-4959/02/$ - see front matter
PII: S1096-4959(01)00514-0

Peterson et al., 1994), Orthoptera (Sakal et al., 1989) and Thysanura (Zinkler and Polzer, 1992).

Tarnished plant bugs, *Lygus lineolaris* (Palisot de Beauvois), are serious pests of cotton, cereals, fruits, vegetables, and alfalfa (Snodgrass et al., 1984; Butts and Lamb, 1991). Management of this important agricultural pest will be aided by an in-depth understanding of the biochemical feeding adaptations of this species. Although tarnished plant bugs are highly efficient in the digestion and absorption of their foods (Cohen, 2000a), the specific properties of proteinases in the salivary glands of *L. lineolaris* are unknown. Therefore, this study was conducted to determine if trypsin-like enzymes are presented in the salivary glands complex of *L. lineolaris*, and to characterize the enzyme at the biochemical and molecular level.

2. Materials and methods

2.1. Insects and sample preparation

Lygus lineolaris adults were obtained from a laboratory culture at the Biological Control and Mass Rearing Research Unit, USDA/ARS, Mississippi. Cultures were reared on artificial diet (Cohen, 2000b) and maintained at 27 ± 1 °C, RH 60 ± 5% with a 14 h light/10 h dark cycle. Voucher specimens were placed in the Mississippi Entomological Museum (No. 72.4), Mississippi State University, Mississippi State, MS, USA. Salivary gland tissue samples were prepared by the method of Cohen (1993) with slight modifications. Young adults (1 week after eclosion) were held for 24 h without food in paper cartons with wet sponges placed on organdie lids. Salivary gland complexes (SGC) from 600 individuals were dissected under light microscopy. Pooled samples of 200 pairs of SGC per replicate were collected, ground in 1 ml of cold (2–4 °C), sterilized distilled water in a small homogenizer and centrifuged at $10\,000 \times g$ for 15 min at 4 °C. The supernatant was transferred to a fresh, cold centrifuge tube, stored at −20 °C and used within 2 weeks. Protein concentration was determined by the BCA protein assay (Pierce, Rockford, IL) using bovine serum albumin (BSA) as the standard.

2.2. Enzyme assay

Trypsin-like-enzyme activity in the samples was assayed using a slight modification method of Stewart (1973). Briefly, the substrate solution was prepared by dissolving N_2-benzoyl-L-arginine-*p*-nitroanilide (L-BA*p*NA) (B-3133, Sigma, St. Louis, MO) in 5 ml of dimethyl methyl sulfoxide (DMSO) before adding 95 ml of 0.05 M Tris–HCl buffer (pH 8.2). Trypsin-like activity was monitored as L-BA*p*NA hydrolysis by adding 100 μl of extract (diluted to 100 μl with buffer for 10 μl of crude sample before the assay) to a well in an assay plate, and adding 100 μl of the substrate solution. The final L-BA*p*NA concentration was 0.5 mg/ml. The absorbance at 420 nm after a 10-min incubation at 37 °C, which is proportional to trypsin-like protease activity, was measured with a plate reader (SPECTRA MAX® Plus, Molecular Devices Corporation, Sunnyvale, CA). The reading from a control (substrate and buffer only) was used to correct for spontaneous hydrolysis of L-BApNA. The specific activity was determined by the method of Asgeirsson et al. (1989).

2.3. Isoelectric focusing

The isoelectric focusing (IEF) method followed Egen et al. (1988) using a Rotofor® Cell System (Bio-Rad Laboratories, Hercules, CA). Before loading the ampholyte solution, the Rotofor® cell was run with distilled water twice at low power (2 W). The preparative cell was equilibrated for 1 h with a 5% ampholyte (pH 3/10) solution at constant power (12 W). One milliliter of SGC extract (protein concentration estimated to be ~2 mg/ml) from *L. lineolaris* adults was added to the equilibrated Rotofor® cell and allowed to focus for 4 h at constant power (12 W). The contents of the focusing chamber were harvested into 20 fractions after isoelectric focusing, and the pH of each fraction was measured. Each fraction was concentrated and washed three times using phosphate buffer in a Centricon-10 (Amicon, Inc., Beverly, MA) at 4 °C and tested for trypsin-like enzyme activity using L-BA*p*NA as described above. The isoelectric point (p*I*) of the trypsin-like enzymes was considered to be the pH of the fraction which exhibited the highest activity.

2.4. Gel filtration

Samples (0.25 ml) of the IEF fraction with the highest activity were applied to a high performance gel filtration column (Superdex® 200 HR 10/30; Pharmacia Biotech Inc., Piscataway, NJ), connect-

ed to a Biologic HR Chromatography System (Bio-Rad Lab., Hercules, CA). The column (bed dimensions is 10×300–310 mm) was equilibrated with elution buffer (phosphate buffer with 0.15 M NaCl, pH 7.4) before injection of sample. The void volume was determined with blue dextran 2000. The elution was conducted at 4 °C, and monitored continuously at 280 nm. The flow rate was 0.5 ml/min and a total of 84 fractions (28 for each replicate) were collected at 1-ml intervals. Trypsin-like enzyme activity was measured for each fraction as described above. For molecular weight estimation, the standards used were bovine serum albumin (66 000), carbonic anhydrase (29 000), and bovine trypsin (23 000). The elution parameter (K_{av}) was calculated as the following: $K_{av}=(V_e-V_0)/(V_t-V_0)$. Where V_e is elution volume for the protein; V_0 is column void volume (elution volume for Blue Dextran 2000); V_t is total bed volume. The standard curve was constructed as the relationship between the elution parameter, K_{av} and the logarithm of standards' respective molecular weights. The molecular weight of the trypsin-like enzymes from *L. lineolaris* was estimated by locating the point on the standard curve that corresponds to the K_{av} value from the most active fraction eluted from the gel filtration column.

2.5. Zymogram gel analysis

The crude and partially purified samples isolated by IEF were further characterized on casein-zymogram gels and incubation with the serine proteinase inhibitors, phenylmethylsulfonyl fluoride (PMSF) (P-7626, Sigma, St. Louis, MO), lima bean trypsin inhibitor (T-9378, Sigma, St. Louis, MO) and chicken egg white trypsin inhibitor (T-2011, Sigma, St. Louis, MO). The crude and the IEF purified extracts not denatured were mixed with an equal volume of zymogram sample buffer (62.5 mM Tris–Cl, pH 6.8, 4% SDS, 25% glycerol and 0.01% Bromophenol Blue). The samples (1–15 μg) were loaded on each well of the opposite sides of a 12% zymogram ready gel (Cat. # 161-1186, Bio-Rad Lab., Hercules, CA). Bovine trypsin (1.5 μg) (T-8003, Sigma, St. Louis, MO), or sample buffer was also loaded to a well for comparison. The gel was run at constant voltage (100 V) with 1× Tris/Glycine/SDS buffer [25 mM Tris, 192 mM glycine and 0.1% (w/v) SDS, pH 8.3]. After electrophoresis, the gel was rinsed in distilled water and incubated in 2.5% Triton X-100 for 30 min at room temperature to remove the SDS and renature the enzymes. Then the gel was rinsed in distilled water, and cut in half to yield two identical gels. One gel was incubated in development buffer (50 mM Tris–Cl, 200 mM NaCl, 5 mM $CaCl_2$ and 0.02% Brij-35, pH 7.8) for 4 h at 37 °C without any inhibitor. The other was incubated under the same conditions, except that inhibitors were added in the development buffer. The final concentrations of inhibitors were 2.5 mM for PMSF, and 0.25% for inhibitors from lima beans and chicken egg white. The gels were rinsed with distilled water after incubation, stained with Coomassie gel stain solution (1 g Coomassie Blue R-250, 450 ml methanol, 450 ml H_2O and 100 ml glacial acetic acid) for 2 h, and then destained in Coomassie gel destain solution (100 ml methanol, 800 ml H_2O and 100 ml glacial acetic acid) overnight.

2.6. pH optimum

Trypsin-like enzyme activity in the IEF fraction with the highest activity was assayed over a pH range of 2–12 at 1 pH unit increments to determine the pH optima of the trypsin-like enzymes in the salivary glands of *L. lineolaris*. For the assay, 10 μl of the collected fraction was mixed with 90 μl of Tris–HCl solution at different pH, and incubated for 10 min at 37 °C in a well of an assay plate. The reaction was started by adding 100 μl of substrate (L-BA*p*NA). The absorbance at 420 nm after incubation for 10 min at 37 °C was measured for three replicates using a plate reader. The blanks (substrate and buffer solution only) were used at each pH to correct for spontaneous hydrolysis. The L-BA*p*NA relative activity at each pH was expressed as a percentage of the pH resulting in the highest activity.

2.7. Cloning trypsin-like protein cDNA

Initially a partial sequence of trypsin-like cDNA in *L. lineolaris* was obtained. Five adults of *L. lineolaris* were ground in TRIzol reagent (Gibco BRL Life-Technologies, Gaithersburg, MD), and total RNA was precipitated with isopropanol. Reverse transcription (RT) was conducted using a

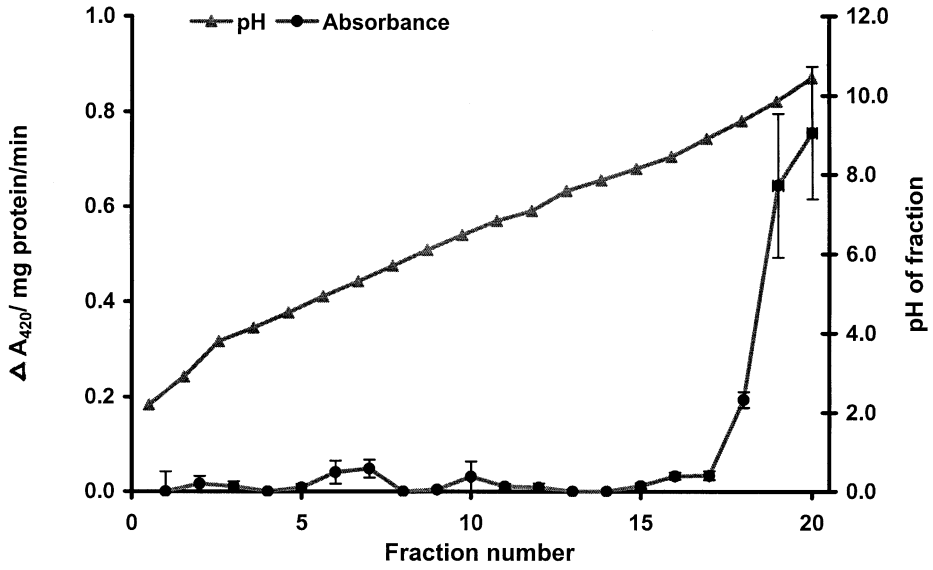

Fig. 1. Isoelectric focusing of trypsin-like enzymes in salivary gland extract of *Lygus lineolaris*, showing pH and the L-BA*p*NAase activity per mg protein for each fraction (mean ± S.E., $n = 3$).

SuperScript Preamplification System for First-Strand cDNA system (Gibco BRL Life-Technologies). The cDNA fragments were amplified using polymerase chain reaction (PCR) with a forward degenerate primer, 5′-CARMGN-ATHGTNGGNGG-3′ (R = A or G, M = A or C, N = A, C, G, or T, H = A, C, or T), and a reverse degenerate primer, 5′-ARNGGNCCNCCNS WRTGNCCYTGRCA-3′ (R = A or G, N = A, C, G, or T, S = G or C, W = A or T, Y = C or T). The primers were designed from two highly conserved regions of serine proteinases (QRIVGG and CQGDSGGPL, respectively) present in both *Manduca sexta* trypsin and chymotrypsin cDNAs (Peterson et al., 1994, 1995). The PCR-amplified DNA fragments (~600 bp) were cloned into a pGEM-T vector (Promega, Madison, WI). The sequence of the clone was determined using an automated sequencer (Model 3700, ABI PRISM®, Foster, CA). The cDNA sequence of the trypsin-like protein was confirmed by a homology search of GenBank (National Center for Biotechnology Information) using the Blastx protocol (Altschul et al., 1990; Gish and States, 1993).

After obtaining the partial sequence of trypsinlike cDNA, two reverse primers, LTR1 (CAAGTCTGCACCTACACGG) and LTR2 (AAGCTAACCTCAACATCGTCG), were designed from the sequence. Reverse transcription was performed using reverse primer LTR1, and the 5′-end of the trypsin precursor cDNA was isolated and C-tailed using the 5′-RACE (rapid amplification of cDNA end) system (Gibco BRL Life-Technologies). Subsequent semi-nested amplification was performed using a forward abridged anchor primer (5′-GGCCACGCGTC GACTAGTACGGGGGGGGGG-3′) and the specific reverse primer LTR2. The DNA fragment resolved from 5′-RACE amplification was cloned into a pGEM-T vector and sequenced from both directions.

The full-length trypsin cDNA was obtained from salivary glands of 100 *L. lineolaris*. Salivary gland RNA was extracted using TRIzol reagent, and subsequently was used as the template for reverse transcription. The resulting cDNA was used for PCR amplification of the full-length cDNA sequence using an oligo-dT reverse primer and one specific forward primer, LTF1 (ACAGCA-TGCAATTAACAACAGT). The LTF1 primer annealed partially to 5′ untranslated region and partially to 5′ coding region of the trypsin cDNA. The amplified cDNA was sequenced from both directions. Sequence analysis tools of the ExPASy Molecular Biology Server of Swiss Institute of Bioinformatics (http://www.expasy.ch/tools/) were used to analyze the deduced protein sequence.

3. Results

3.1. Isoelectric focusing

Twenty fractions were harvested from preparative IEF. The highest L-BA*p*NA hydrolysis activity from crude SGC extract of *L. lineolaris* was found in fraction number 20 (Fig. 1), which had a pH of 10.4. The specific activity of BA*p*NAase in this fraction was much higher than that of the crude SGC extract, the IEF resulted in 23-fold purification of BA*p*NAase. There were several minor peaks that showed slight L-BA*p*NA hydrolysis activity before the major peak.

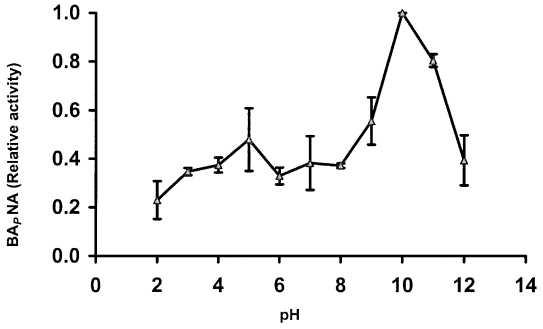

Fig. 3. L-BA*p*NAase activity of the IEF purified salivary gland extract of *Lygus lineolaris* at different pH (values are means ± S.E., $n = 3$, and expressed relative to maximal activity).

3.2. Gel filtration

L-BA*p*NA hydrolysis activity was found predominantly in fraction 19 from the calibrated Superdex® 200 column. Based on elution of the protein standards, this fraction eluted with an apparent molecular weight of 23,000.

3.3. Zymogram gel analysis

Zymogram analysis revealed a single caseinolytic band for the bovine trypsin (lane 3), and revealed one strong band and a faint band for the fraction isolated from the SGC extract (lane 4) by preparative IEF (Fig. 2a). The strong activity band of caseinolytic activity of *L. lineolaris* migrated slightly slower than that of bovine trypsin. In addition to this band, several other caseinolytic activity bands were detected in lane 2 which was loaded with crude SGC extracts, and lane 4 which contained the IEF-purified SGC extracts. Lima bean trypsin inhibitor completely suppressed the caseinolytic activity of bovine trypsin, the IEF-purified extracts, and crude SGC extracts on the zymogram gel (gel was blank and not shown). Also, incubation with PMSF completely inhibited bovine trypsin and caseinolytic activity in the IEF-purified SGC extract of the *L. lineolaris* (gel not shown). However, the inhibitor from chicken egg white only partially inhibited protease activity on the zymogram gel (Fig. 2b) and protease activity

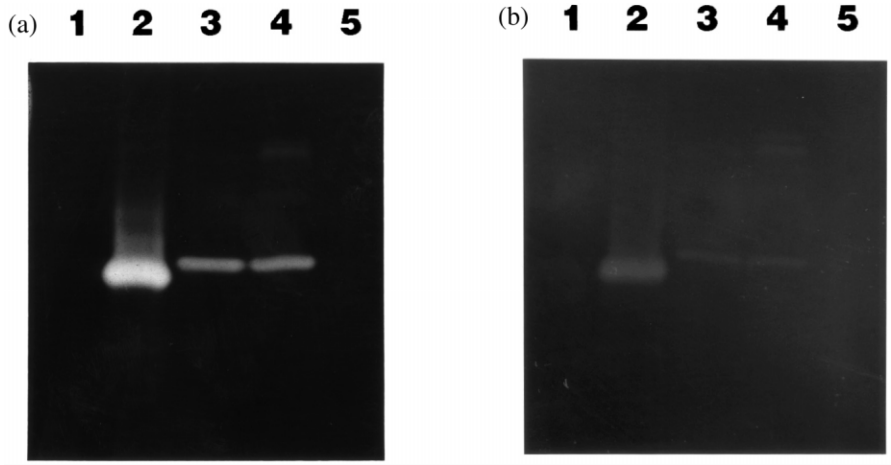

Fig. 2. Zymogram analysis and the effects of serine proteinase inhibitors on caseinolytic activity of enzymes in salivary glands of *Lygus lineolaris* on 12% gels containing casein. Lane 1, sample buffer only; lane 2, crude SGC extract of *L. lineolaris*; lane 3, bovine trypsin; lane 4, the IEF purified SGC extract of *L. lineolaris*; and lane 5, sample buffer only. (a) Gel was incubated without inhibitor, control; (b) gel was incubated with inhibitor from chicken egg white.

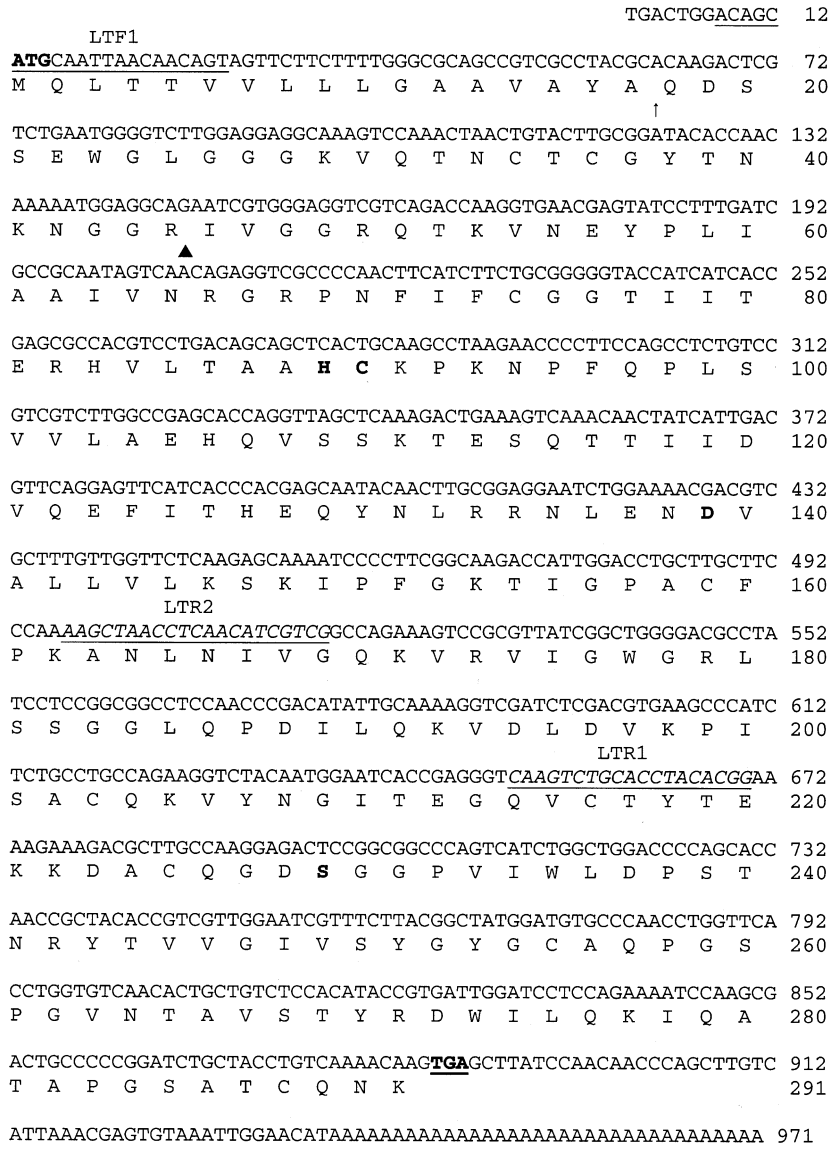

Fig. 4. Nucleotide and deduced amino acid sequences of trypsin-like precursor isolated from *L. lineolaris* RNA. ATG = start codon; TGA = termination codon; ATTAAA = potential polyadenylation signal; ↑ = predicted signal peptide cleavage site; ▲ = predicted activation peptide cleavage site. IVGG are conserved N-terminal residues. Primer sequences used for cloning and sequencing cDNA are underlined and labeled on the top. Complimentary sequences of the reverse primers are indicated with italic case. The sequence has been deposited in GenBank with accession number AF414430.

bands were apparent in lane 3 (bovine trypsin), lane 4 (the IEF- purified SGC extract) and lane 2 (SGC crude extract).

3.4. pH optimum

Trypsin-like activity from the IEF-purified SGC extract of *L. lineolaris* was detected over a broad pH range. Maximum activity was measured at pH 10 (Fig. 3). A small peak at pH 5 was noted before the major peak of trypsin-like enzymes.

3.5. Trypsin precursor cDNA

Three independent clones of trypsin precursor cDNA from the salivary glands had identical sequences. The cDNA sequences were subjected to a GenBank homology search, and the results

suggested that the cDNA encoded a trypsin-like precursor. The cloned trypsin precursor cDNA contained 971 nucleotides and included the start codon ATG at positions 13–15, the termination codon TGA at positions 886–888, and a potential polyadenylation signal, ATTAAA, at positions 913–918 (Fig. 4). The open reading frame consisted of 873 nucleotides.

The deduced trypsin precursor contained 291 amino acid residues. A 17-residue signal peptide was predicted by using SignalP program (Nielsen et al., 1997). The proenzyme for the trypsin-like protein contained 28 amino acid residue-activation peptides, which ended in a basic arginine residue, suggesting that tryptic cleavage (Arg-Ile or Lys-Ile bond) might be involved in activation. The calculated molecular masses for the trypsin precursor and predicted mature enzyme were 31.3 kDa, and 26.7 kDa, respectively, with 18 negatively charged residues (Asp + Glu) and 26 positively charged residues (Arg + Lys) in the mature form. The calculated p*I* was 9.09 for the mature enzyme. This predicted p*I* was similar to the experimentally observed p*I*.

4. Discussion of results

4.1. Isoelectric point (pI)

The p*I* value (10.4) for trypsin-like enzymes from the salivary gland of *L. lineolaris* is similar to that (10.3) of *Choristoneura fumiferana* larva (Milne and Kaplan, 1993). However, the p*I* value of *L. lineolaris* was much more alkaline compared with those (p*I* 3.5–4.6) of trypsin from *Aedes aegypti*, *Locusta migratoria*, *Pheropsophus aequinoctialis*, and *Anopheles quadrimaculatus* (Graf and Briegel, 1985; Ferreira and Terra, 1989; Sakal et al., 1989; Graf et al., 1991). These differences may reflect the phylogenetic relationships, different food resources, or differences between enzymes from the salivary gland and those that originate from the midgut.

4.2. Molecular weight

The molecular weight of trypsin-like protein in the salivary gland of *L. lineolaris* estimated by gel filtration was 23,000. The result is very similar to those reported for *Costelytra zealandica* (23,000) (Christeller et al., 1989); *Bombyx mori* (22,000) (Sasaki and Suzuki, 1982); one (23,000) of the three trypsins from the larval gut of *Spodoptera litura* (Ahmad et al., 1980); *Locusta migratoria*, (24,000) (Sakal et al., 1989); *Manduca sexta* (24,000) (Miller et al., 1974); and *Helicoverpa armigera* (24,000) (Johnston et al., 1991); one (24,000) of the two trypsins isolated from the larval midgut of *Sesamia nonagrioides* (Novillo et al., 1999). However, the molecular weight of trypsin-like protein in *L. lineolaris* salivary glands is much different from the trypsin (13,000) from the midgut epithelium of *B. mori* (Eguchi and Kuriyama, 1985), and one (53,000) of the three trypsins from *S. litura* (Ahmad et al., 1980). These differences may reflect the phylogenetic relationships, or the origin of the enzymes, i.e. gut or salivary glands.

4.3. pH optimum

The alkaline pH optimum for the SGC extract of *L. lineolaris* is in agreement with previous reports of heteropteran insects (Laurema et al., 1985; Cohen, 1993; Colebatch et al., 2001). In addition, Colebatch et al. (2001) reported that the highest level of proteinase activity detected in saliva of the green mirid, *Creontiades dilutus*, was at pH 10. Similarly, an alkaline pH optimum for midgut trypsin-like enzymes from the lesser grain borer, *Rhyzopertha dominica*, was found by Zhu and Baker (1999).

In our study, one peak of proteinase activity was detected before the major activity peak of the trypsin-like enzymes of *L. lineolaris* SGC extract. This may suggest that the SGC extract of *L. lineolaris* contains other proteinase because hydrolysis of BA*p*NA could be due to either trypsin-like or cathepsin-like enzymes. Houseman et al. (1984) studied the effect of pH on the hydrolysis of BA*p*NA by *Phymata wolffii* Stäl gut extracts using phosphate buffer, and found that the pH value for the maximal hydrolysis was 5.5 for cathepsin B, a cysteine proteinase. The small peak occurring before the major trypsin-like enzyme peak was close to the pH value of 5.5 and this might imply that the SGC extract of *L. lineolaris* contains a cysteine protease. The putative cysteine protease may be of lysosomal origin appearing as an artifact resulting from maceration of salivary gland tissue. Whether cysteine protease genes are expressed in the salivary glands of *L. lineolaris* and contribute significantly to the digestive process remains to be determined.

4.4. Zymogram gel analysis

Interestingly, trypsin-like enzymes and other proteinases from *L. lineolaris* were more sensitive to the plant-derived protein inhibitor (lima bean trypsin inhibitor) than the animal-derived inhibitor (chicken egg white trypsin inhibitor). This result is consistent with observations by Purcell et al. (1992) who found that plant derived proteinase inhibitor, soybean trypsin inhibitor, was more effective than ovomucoid (inhibitor from chicken egg white) against proteases in midguts from three lepidopterans and the boll weevil.

Novillo et al. (1999) also reported that lima bean inhibitor inhibits one of two trypsins isolated from the larval midgut of *S. nonagrioides*. Richardson (1977) pointed out that inhibitors of serine proteases may be active against enzymes in other groups. For instance, an inhibitor from broad beans inhibited both serine and thiol proteases (Warsy et al., 1974). In this study the lima bean trypsin inhibitor and the low molecular weight inhibitor, PMSF, not only suppressed trypsin-like enzymes, but also other enzyme activities in the SGC crude extracts of *L. lineolaris*. Multiple protease activity bands were detected by zymogram gel analysis with crude SGC extract of *L. lineolaris.* Similarly, other researchers have also observed multiple activity bands in salivary gland extract of the green mirid, *Creontiades dilutus*, (Colebatch, 1999) and midgut extract of the lesser grain borer, *R. dominica,* (Zhu and Baker, 1999) by using zymogram gels.

4.5. Trypsin-like protein cDNA

A scan of the PROSITE database revealed that the deduced protein sequence belongs to the trypsin family of serine proteases which contains a conserved histidine active site sequence pattern [LIVM]-[ST]-A-[STAG]-H-C and a conserved serine active site sequence pattern [DNSTAGC]-[GSTAPIMVQH]-x(2)-G-[DE]-S-G-[GS]-[SAPHV]-[LIVMFYWH]-[LIVMFYS-TANQH]. Amino acid residues 74–264 were predicted to be a catalytic domain for trypsin-like serine proteinases.

The predicted amino acid sequence encoded by the *L. lineolaris* cDNA was aligned with four highly homologous insect trypsin-like proteases from GenBank (Gap open = 3, gap extension = 1, Fig. 5). This sequence contained all of the conserved residues in the putative active site, His^{89}, Asp^{139}, Ser^{229}, which forms the catalytic triad in serine proteases (Kraut, 1977; Wang et al., 1993; Peterson et al., 1994). The sequence IVGG at positions 46–49 is highly conserved in many trypsin- and chymotrypsin-like proteinases and marks the N-termini of the active enzymes (Davis et al., 1985; Wang et al., 1993). Six conserved cysteine residues, predicted to occur in disulfide bridge configurations among trypsins and chymotrypsins, were located at positions 74, 90, 203, 216, 225, and 255. The cDNA of *L. lineolaris* encoded an extra pair of cysteine residues at positions 159 and 288, which are present in three other insect proteinases; however, the pairing mode of cysteine residues for disulfide bridge configurations in this insect has not been established. It is very likely that *L. lineolaris* trypsin has free cysteine residues, and these free cysteines in the active site or elsewhere on the protein make the enzyme sensitive to some typical thiol-reacting inhibitors (Jongsma et al., 1996; Gatehouse et al., 1997).

The residues (Asp^{223}, Gly^{252}, Gly^{262}), which define the substrate binding pocket were highly conserved in all of these trypsin-like enzymes. The Asp^{223} has been predicted to be located at the bottom of the binding pocket, which determines specificity in both invertebrate and vertebrate trypsins by interacting through ionic forces with a lysine or arginine residue at the substrate cleavage site (Graf et al., 1988; Hedstrom et al., 1992; Perona and Craik, 1995; Wang et al., 1993).

A search of the GenBank database using the BLASTP program revealed that the deduced protein sequence of *L. lineolaris* was most similar to proteins in the serine proteinase trypsin/chymotrypsin family. Some of the best matches included the uncharacterized *Drosophila* genes. The most similar sequence was that of a protease in the African malaria mosquito *Anopheles gambiae* (GenBank CAA89967, Siden-Kiamos et al., 1996). Other highly similar insect trypsin-like sequences included serine proteinase from the flesh fly, *Sarcophaga peregrina*, (GenBank BAA22400, Nakajima et al., 1997), serine proteinase precursor from the African malaria mosquito *A. gambiae* (GenBank S55493, direct submission by Siden-Kiamos et al., 1996), and trypsin RdoT1 precursor from the lesser grain borer, *Rhyzopertha dominica*, (GenBank AAD31267, Zhu and Baker, 1999). Sequence alignment using CLUSTALW indicated

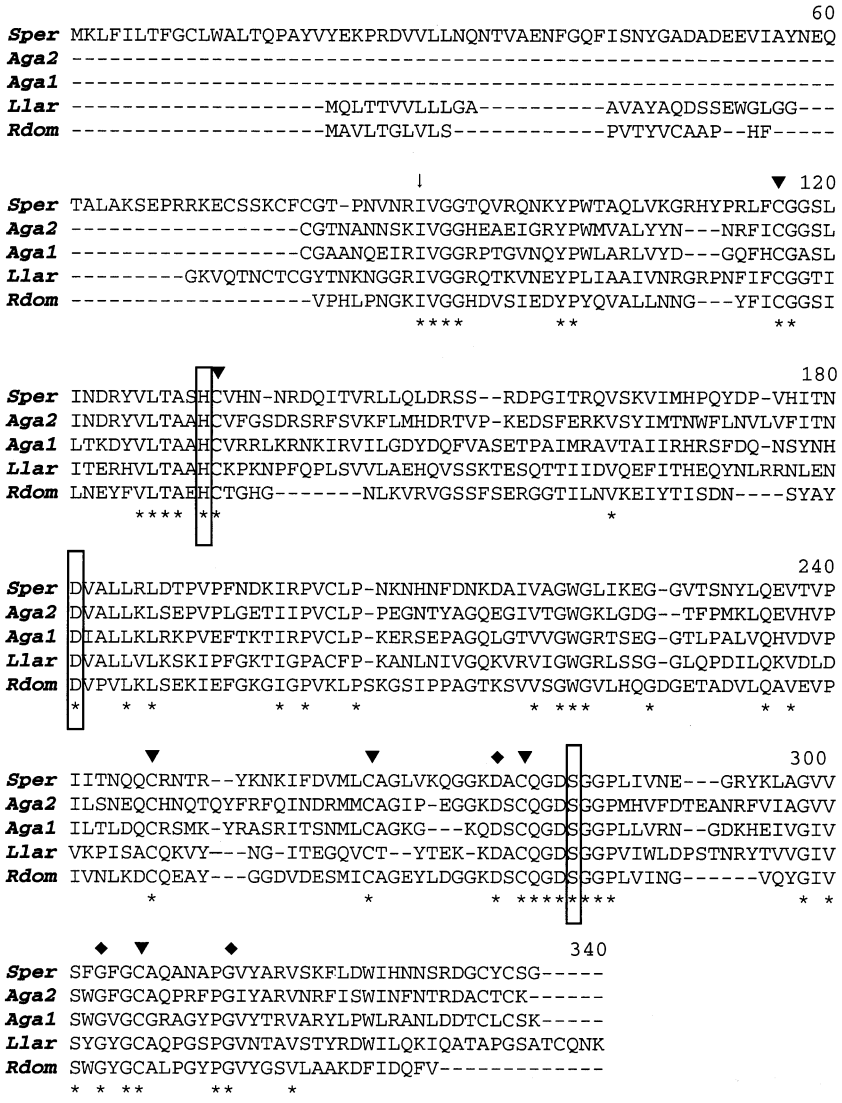

Fig. 5. Predicted amino acid sequence of trypsin-like precursor from *L. lineolaris* and alignment with four other insect trypsin-like sequences. *Sper*=protease from *S. peregrina*, *Aga1*=serine proteinase from *A. gambiae*, *Aga2*=serine proteinase precursor from *A. gambiae*, *Llar*=deduced trypsin precursor from *L. lineolaris*, and *Rdom*=trypsin RdoT1 precursor from *R. dominica*. Functionally important residues His^{89}, Asp^{139}, and Ser^{229} are indicated by boxed letters. Cysteines corresponding to the sites of predicted disulfide bridges are marked with solid triangle (▼) on the top. Conserved residues for the trypsin binding pocket, Asp^{223}, Gly^{252}, and Gly^{262}, are indicated by (◆) at the top of the sequences. Identical residues among all five sequences are indicated with stars (*) at the bottom of the sequences. The arrow (↓) indicates the N-terminal residues of the active enzymes. Hyphens represent sequence alignment gaps.

that trypsin-like protein from *L. lineolaris* shared 30–33% sequence identity with those four insect serine proteinases.

The presence of trypsin-like enzyme and cDNA in the salivary glands indicated that *L. lineolaris* has evolved or retained trypsin-like genes involved in protein digestion in this tissue. GenBank homology search and CLUSTALW multiple sequence alignment confirmed that cloned cDNA sequence from *L. lineolaris* salivary gland encodes a trypsin-like enzyme protein. Molecular cloning using general primers for trypsin, and chymotrypsin genes exclusively generated trypsin-like cDNA clones (data not shown), which also suggested that trypsin-like mRNAs are dominant transcripts within the salivary glands of *L. lineolaris*. The knowledge gained from this study may be helpful in designing new strategies, such as using digestive enzyme

inhibitor to reduce digestive capacities and therefore enhance efforts to manage this economically important pest.

Acknowledgments

We thank Patrick Crittenden for technical help, Gay McCain and Brenda Woods for rearing insects, Drs Randal Buddington, Clarence Collison and Dawn Luthe for reviewing an earlier version of this manuscript. We also thank Drs Janet Alverson and Frank Callahan for use of their laboratory facilities, Drs Din-Pow Ma, Wai K. Ma and Tibor Pechan for their helpful suggestions. Approved for publication as Journal Article No. BC9903 of the Mississippi Agricultural and Forestry Experiment Station, Mississippi State University.

References

Agusti, N., Cohen, A.C., 2000. *Lygus hesperus* and *L. lineolaris* (Hemiptera: Miridae), phytophages, zoophages, or omnivores: evidence of feeding adaptations suggested by the salivary and midgut digestive enzymes. J. Entomol. Sci. 35, 176–186.

Ahmad, Z., Saleemuddin, M., Siddi, M., 1980. Purification and characterization of three alkaline proteases from the gut of the larva of armyworm, *Spodoptera litura*. Insect Biochem. 10, 667–673.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Asgeirsson, B., Fox, J.W., Bjarnason, B.J., 1989. Purification and characterization of trypsin from the poikilotherm *Gadus morhua*. Eur. J. Biochem. 180, 85–94.

Baker, J.E., 1981. Resolution and partial characterization of the digestive proteinases from larvae of the black carpet beetle. In: Bhaskaran, G., Friedman, S., Rodriguez, J.G. (Eds.), Current Topics in Insect Endocrinology and Nutrition, Plenum Press, New York, pp. 283–315.

Butts, R.A., Lamb, R.J., 1991. Seasonal abundance of three *Lygus* species (Heteroptera: Miridae) in oilseed rape and alfalfa in Alberta. J. Econ. Entomol. 84, 450–456.

Chapman, R.F., 1998. The Insects: Structure and Function. Cambridge University Press, New York 770 pp.

Christeller, J.T., Shaw, B.D., Gradiner, S.E., Dymock, J., 1989. Partial purification and characterization of the major midgut proteases of grass grub larvae (*Costelytra zealandica*), Coleoptera: Scarabaeidae. Insect Biochem. 19, 221–231.

Cohen, A.C., 1993. Organization of digestion and preliminary characterization of salivary trypsin-like enzymes in a predaceous heteropteran, *Zelus renardii*. J. Insect Physiol. 39, 823–829.

Cohen, A.C., 1998. Solid-to-liquid feeding: the inside(s) story of extra-oral digestion in predaceous Arthropoda. Am. Entomologist 44, 103–117.

Cohen, A.C., 2000a. A review of feeding studies of *Lygus* spp. with emphasis on artificial diets. Southwestern Entomologist 23, 111–119.

Cohen, A.C., 2000b. New oligidic production diet for *Lygus hesperus* Knight and *L. lineolaris* (Palisot de Beauvois). J. Entomol. Sci. 35, 301–310.

Colebatch, G.M., 1999. Organisation of Digestion and Digestive Proteinases in the Green Mirid, *Creontiades dilutus*, Ph.D. Thesis, Australian National University, 184 pp (submitted).

Colebatch, G.M., East, P., Cooper, P., 2001. Preliminary characterization of digestive proteases of the green mirid *Creontiades dilutus* (Hemiptera: Miridae). Insect Biochem. Molec. Biol. 31, 415–423.

Davis, C.A., Riddell, D.C., Higgins, M.J., Holden, J.J.A., White, B.N., 1985. A gene family in *Drosophila melanogaster* coding for trypsin-like enzymes. Nucleic Acids Res. 13, 6605–6619.

Egen, N.B., Bliss, M., Mayersohn, M., Owens, S.M., Arnold, L., Bier, M., 1988. Isolation of monoclonal antibodies to phencyclidine from ascites fluid by preparative isoelectric focusing in the Rotofor. Anal. Biochem. 172, 488–494.

Eguchi, M., Kuriyama, K., 1985. Purification and characterization of membrane-bound alkaline proteases from the midgut tissue of the silk-worm *Bombyx mori*. J. Biochem. 97, 1437–1445.

Ferreira, C., Terra, W.R., 1989. Spatial organization of digestion, secretory mechanisms and digestive enzyme properties in *Pheropsophus aequinoctialis* (Coleoptera: Carabidae). Insect Biochem. 19, 383–391.

Gatehouse, L.N., Shannon, A.L., Burgess, E.P.J., Christeller, J.T., 1997. Characterization of major midgut proteinase cDNAs from *Helicoverpa armigera* larvae and changes in gene expression in response to four proteinase inhibitors in the diet. Insect Biochem. Molec. Biol. 27, 929–944.

Giebel, W., Zwilling, R., Pfleiderer, G., 1971. The evolution of endopeptidases — XII. The proteolytic enzymes of the honeybee (*Apis mellifica* L.). Comp. Biochem. Physiol. B 38, 197–210.

Gish, W., States, D.J., 1993. Identification of protein coding regions by database similarity search. Nat. Genet. 3, 266–272.

Goodchild, A.J.P., 1952. A study of the digestive system of the West African cacao capsid bugs (Hemiptera Miridae). Proc. Zool. Soc. Lond. 122, 543–572.

Gooding, R.H., Huang, C.T., 1969. Trypsin and chymotrypsin from the beetle *Pterostichus melanarius*. J. Insect Physiol. 15, 325–339.

Gooding, R.H., Rolseth, B.M., 1976. Digestive processes of haematophagous insects. XI. Partial purification and some properties of six proteolytic enzymes from the tsetse fly *Glossina morsitans* Westwood (Diptera: Glossinidae). Can. J. Zool. 54, 1950–1959.

Graf, R., Briegel, H., 1985. Isolation of trypsin isozymes from the mosquito *Aedes aegypti* L. Insect Biochem. 15, 611–618.

Graf, L., Jancso, A., Szilagyi, L., Hegyi, G., Pinter, K., Naray-Szabo, G., Hepp, J., Medzihradszky, K., Rutter, W.J., 1988. Electrostatic complementarity within the substrate-binding pocket of trypsin. Proc. Natl. Acad. Sci. 85, 4961–4965.

Graf, R., Boehlen, P., Briegel, H., 1991. Structural diversity of trypsin from different mosquito species feeding on vertebrate blood. Experientia 47, 603–609.

Hedstrom, L., Szilagyi, L., Rutter, W.J., 1992. Converting trypsin to chymotrypsin: the role of surface loops. Science 255, 1249–1253.

Hori, K., 1970. Some variations in the activities of salivary amylase and protease of *Lygus disponsi* Linnavuori (Hemiptera: Miridae). Appl. Entomol. Zool. 5, 51–61.

Houseman, J.G., Morrison, P.E., Downe, A.E.R., 1984. Cathepsin B and aminopeptidase in the posterior midgut of *Phymata wolffii* Stal (Hemiptera: Phymatidae). Can. J. Zool. 63, 1288–1291.

Johnston, K.A., Lee, M.J., Gatehouse, J.A., Anstee, J.H., 1991. The partial purification and characterization of serine protease activity in midgut of larval *Helicoverpa armigera*. Insect Biochem. 21, 389–397.

Jongsma, M.A., Peters, J., Stiekema, W.J., Bosch, D., 1996. Characterization and partial purification of gut proteinase of *Spodoptera exigua* Hübner (Lepidoptera: Noctuidae). Insect Biochem. Molec. Biol. 26, 185–193.

Kraut, J., 1977. Serine proteases: structure and mechanism of catalysis. Ann. Rev. Biochem. 46, 331–358.

Laurema, S., Varis, A.L., Miettinen, H., 1985. Studies on enzymes in the salivary glands of *Lygus rugulipennis* (Hemiptera Miridae). Insect Biochem. 15, 211–224.

Lemos, F.J.A., Terra, W.R., 1992. Soluble and membrane-bound forms of trypsin-like enzymes in *Musca domestica* larval midguts. Insect Biochem. Molec. Biol. 22, 613–619.

Miller, J.W., Kramer, K.J., Law, J.H., 1974. Isolation and partial characterization of the larval midgut trypsin from the tobacco hornworm, *Manduca sexta*, Johannson (Lepidoptera: Sphingidae). Comp. Biochem. Physiol. B 48, 117–129.

Milne, R., Kaplan, H., 1993. Purification and characterization of a trypsin-like digestive enzyme from spruce budworm (*Choristoneura fumiferana*) responsible for the activation of δ-endotoxin from *Bacillus thuringiensis*. Insect Biochem. Molec. Biol. 23, 663–673.

Nakajima, Y., Tsuji, Y., Homma, K., Natori, S., 1997. A novel protease in the pupal yellow body of *Sarcophaga peregrina* (flesh fly). Its purification and cDNA cloning. J. Biol. Chem. 272, 23805–23810.

Nielsen, H., Engelbrecht, J., Brunak, S., Heijne, G., 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10, 1–6.

Novillo, C., Castáera, P., Ortego, F., 1999. Isolation and characterization of two digestive trypsin-like proteinases from larvae of the stalk corn borer, *Sesamia nonagrioides*. Insect Biochem. Molec. Biol. 29, 177–184.

Perona, J.J., Craik, C.S., 1995. Structural basis of substrate specificity in the serine proteases. Protein Sci. 4, 337–360.

Peterson, A.M., Barillas-Mury, C.V., Wells, M.A., 1994. Sequence of three cDNAs encoding an alkaline midgut trypsin from *Manduca sexta*. Insect Biochem. Molec. Biol. 24, 463–471.

Peterson, A.M., Fernando, G.P., Wells, M.A., 1995. Purification, characterization and cDNA sequence of an alkaline chymotrypsin from the midgut of *Manduca sexta*. Insect Biochem. Molec. Biol. 25, 765–774.

Purcell, J.P., Greenplate, J.T., Sammons, R.D., 1992. Examination of midgut luminal proteinase activities in six economically important insects. Insect Biochem. Molec. Biol. 22, 41–47.

Richardson, M., 1977. The proteinase inhibitors of plants and micro-organisms. Phytochemistry 16, 159–169.

Sakal, E., Applebaum, S.W., Birk, Y., 1989. Purification and characterization of trypsins from the digestive tract of *Locusta migratoria*. Int. J. Pept. Prot. Res. 34, 498–505.

Sasaki, T., Suzuki, Y., 1982. Alkaline proteases in digestive juice of the silkworm, *Bombyx mori*. Biochim Biophys. Acta 703, 1–10.

Schumaker, T.T.S., Cristofoletti, P.T., Terra, W.R., 1993. Properties and compartmentalization of digestive carbohydrases and proteases in *Scaptotrigona bipunctata* (Apidae: Meliponinae) larvae. Apidologie 24, 3–17.

Siden-Kiamos, I., Skavdis, G., Rubio, J., Papaginnakis, G., Louis, C., 1996. Isolation and characterization of three serine protease genes in the mosquito *Anopheles gambiae*. Insect Molec. Biol. 5, 61–71.

Snodgrass, G.L., Scott, W.P., Smith, J.W., 1984. Host plants and seasonal distribution of the tarnished plant bug (Hemiptera: Miridae) in the Delta of Arkansas, Louisiana and Mississippi. Environ. Entomol. 13, 110–116.

Stewart, K.K., 1973. A method for automated analyses of the activities of trypsin, chymotrypsin and their inhibitors. Anal. Biochem. 51, 11–18.

Terra, W.R., Ferreira, C., 1983. Further evidence that enzymes involved in the final stages of digestion by *Rhynchosciara* do not enter the endoperitrophic space. Insect Biochem. 13, 143–150.

Wang, S., Magoulas, C., Hickey, D.A., 1993. Isolation and characterization of a full-length trypsin-encoding cDNA clone from the lepidopteran insect, *Choristoneura fumiferana*. Gene 136, 375–376.

Warsy, A.S., Norton, G., Stein, M., 1974. Protease inhibitors from broadbean; isolation and purification. Phytochemistry 13, 2481–2486.

Zeng, F., Cohen, A.C., 2000. Comparison of alpha-amylase and protease activities of a zoophytophagous and two phytozoophagous heteroptera. Comp Biochem. Physiol. A 126, 101–106.

Zhu, Y.C., Baker, J.E., 1999. Characterization of midgut trypsin-like enzymes and three trypsin cDNAs from the lesser grain borer, *Rhyzopertha dominica* (Coleoptera: Bostrichidae). Insect Biochem. Mol. Biol. 29, 1053–1063.

Zinkler, D., Polzer, M., 1992. Identification and characterization of digestive proteinases from the firebrat *Thermobia domestica*. Comp. Biochem Physiol. B 103, 669–673.